J. N. Am. Benthol. Soc., 2008, 27(4):920–931

DOI: 10.1899/07-152.1
Published online: 28 October 2008

Assessing the attributable risks, relative risks, and regional extents of aquatic stressors

John Van Sickle[1] AND **Steven G. Paulsen**[2]

National Health and Environmental Effects Research Laboratory, US Environmental Protection Agency, 200 SW 35th St., Corvallis, Oregon 97333 USA

Abstract. A major goal of the national aquatic surveys being conducted by the Environmental Protection Agency in partnership with the states and tribes is to assess the relative importance, at a regional scale, of stressors that impact aquatic biota. The Wadeable Streams Assessment (WSA) was a prototype of these surveys, and it assessed 8 individual water-chemistry, physical-habitat, and landuse stressors in 2 ways. First, the WSA estimated the total length of streams in a region that was deemed to be in poor condition for each stressor considered separately. Estimates of stressor *extent* describe the prevalence and regional breadth of each stressor's potential effects. Second, the WSA estimated each stressor's *relative risk*, a measure widely used in human epidemiology. Relative risk estimates a stressor's association with biota in terms of the likelihood that poor stressor conditions and poor biota conditions co-occur in a region's streams. We describe how the *population attributable risk*, also borrowed from epidemiology, combines extent and relative risk into a single, overall measure of a stressor's regional effect. The attributable risk of a stressor is the % reduction in the regional extent of poor biological condition (measured here by the macroinvertebrate index of biotic integrity [MIBI]), that presumably would result from eliminating that stressor. Under the attributable risk assumptions, for example, the WSA data imply that the nationwide extent of poor MIBI conditions would be reduced by an estimated 26% if excess total P were eliminated as a stressor, but only by 3% if excess salinity were eliminated. We also illustrate the attributable risk for the combined effects of multiple, correlated stressors. Last, we discuss how best to interpret and apply attributable risk estimates.

Key words: Wadeable Streams Assessment, stressor ranking, macroinvertebrate, categorical data analysis.

The Wadeable Streams Assessment (WSA) was designed to evaluate the condition of streams and rivers throughout the contiguous US (USEPA 2006). As a critical component of this evaluation, the WSA assessed the extents and probable effects of selected aquatic stressors within major subregions and nationwide. Numerous statistical tools are available to help assess stressors for individual streams or watersheds (USEPA 2000, Yuan and Norton 2004, http://cfpub.epa.gov/caddis), but few of these tools are clearly applicable at broad regional and national scales.

Van Sickle et al. (2006) suggested 2 additional tools that seem particularly useful for regional surveys, such as the WSA. First, they suggested estimating the regional *extent* of each stressor. A WSA-type survey is designed to describe the regional distributions of ≥1 stressor indicators. By specifying a range of elevated values that describe poor condition for any given stressor, one can estimate the percentage of stream length in a region that is in poor condition for that stressor directly from its regional distribution. This percentage expresses the overall extent, or regional prevalence, of elevated stressor levels without attempting to identify specific streams or rivers having those levels. An aquatic stressor in poor condition for a large proportion of regional stream length is clearly of potential concern.

A stressor is also of regional concern if its elevated levels, whenever they do occur, have a strong and deleterious effect on stream ecosystems. Van Sickle et al. (2006) proposed calculating *relative risk*, a measure borrowed from epidemiology, to describe the observed association between poor stressor conditions and poor conditions of a biological indicator such as an index of biological integrity (IBI; Karr 1981). Extent and relative risk estimates can help to evaluate policy and management options, but a single measure that

[1] E-mail addresses: vansickle.john@epa.gov
[2] paulsen.steve@epa.gov

TABLE 1. Condition class thresholds for the macroinvertebrate index of biotic integrity (MIBI) and 7 stressor indicators for 9 aggregated ecoregions (Fig. 1) sampled in the Environmental Protection Agency Wadeable Streams Assessment. In each cell, poor and good conditions are defined by the 1st and 2nd inequalities, respectively. Values between the 2 thresholds are designated as fair condition. Aggregated ecoregions are further grouped into 3 major regions (Plains/Lowland [PLLOW], West [WEST], and Eastern Highlands [EHIGH]) for reporting of stressor and biological conditions. For a unitless indicator, parentheses contain the range of values observed in all WSA samples. N = total N, P = total P, RD = riparian disturbance, SS = stream sediments, I-SFH = instream fish habitat, RVC = riparian vegetation cover, SAL = salinity.

Aggregated ecoregion (major region)	MIBI (0, 94)	N (μg/L)	P (μg/L)	RD (0, 5.9)	SS[a] (−4.4, 1.9)	I-SFH (0, 3.2)	RVC (0, 2.8)	SAL (μS/cm)
Coastal Plains (PLLOW)	<42	>2078	>108	>1.5	<–3.1 or >0.2	<0.41	<0.62	>1000
	56	≤1092	≤56	<0.33	–2.4 and ≤–1.3	0.49	1.02	≤500
Northern Appalachians (EHIGH)	<49	>441	>16	>1.5	<–1.4 or >0.0	<0.25	<0.50	>1000
	63	≤329	≤8	<0.33	–0.9 and ≤–0.3	0.35	0.90	≤500
Northern Plains (PLLOW)	<41	>1570	>183	>1.5	<–2.6 or >0.0	<0.17	<0.27	>2000
	55	≤948	≤92	<0.33	–2.0 and ≤–0.5	0.31	0.49	≤1000
Southern Appalachians (EHIGH)	<37	>535	>24	>1.5	<–1.2 or >0.7	<0.10	<0.36	>1000
	51	≤296	≤18	<0.33	–0.6 and ≤0.2	0.28	0.76	≤500
Southern Plains (PLLOW)	<30	1570	>95	>1.5	<–2.6 or >0.0	<0.17	<0.27	>2000
	44	≤698	≤52	<0.33	–2.0 and ≤–0.5	0.31	0.49	≤1000
Temperate Plains (PLLOW)	<31	>3210	>338	>1.5	<-2.6 or >0.0	<0.18	<0.21	>2000
	45	≤1750	≤165	<0.33	–2.0 and ≤–0.5	0.25	0.35	≤1000
Upper Midwest (PLLOW)	<34	>1300	>45	>1.5	<–1.5 or >0.2	<0.15	<0.27	>1000
	48	≤716	≤22	<0.33	–1.3 and ≤–0.5	0.65	0.88	≤500
Western Mountains[b] (WEST)	<40	>229	>36	>1.5			<0.23	>1000
	54	≤131	≤14	<0.33			0.67	≤500
Northern Rockies					<–1.8 or >0.1	<0.18		
					–1.1 and ≤–0.4	0.34		
Pacific Northwest					<–1.3 or >0.6	<0.14		
					–0.7 and ≤0.1	0.33		
Southern Rockies					<–1.6 or >0.3	<0.10		
					–0.9 and ≤–0.2	0.37		
Southwest					<–1.3 or >0.3	<0.15		
					–0.6 and ≤0.1	0.65		
Xeric (WEST)	<40	>462	>70	>1.5	<–1.7 or >0.3	<0.13	<0.32	>1000
	53	≤246	≤36	<0.33	–0.9 and ≤–0.1	0.27	0.60	≤500

[a] Any value that lies beyond the stated thresholds in either direction is assigned poor condition, and any value within the stated interval is assigned good condition; other values are designated as fair condition.
[b] Different SS and I-SFH thresholds are specified for each of 4 subecoregions.

combines information from the 2 estimates would be helpful.

We propose using *population attributable risk*, also borrowed from epidemiology, to combine relative risk and extent into an estimate of a stressor's net effect on a regional population of streams. We estimate the population attributable risks of 7 WSA stressors, using the macroinvertebrate IBI (MIBI) as a biological response indicator. We also illustrate how to estimate the combined attributable risk of the joint effects of multiple correlated stressors. Before discussing attributable risk, we review the definitions of stressor extent and relative risk, and we describe how these measures were applied to 7 stressors in the WSA, as summarized by Paulsen et al. (2008). We use the extent and relative risk results to motivate development of attributable risk. We use the attributable risk results for individual stressors, in turn, to motivate an assessment of the combined attributable risk for closely related stressors. Therefore, we have structured our paper as a sequence of methods, followed immediately by results, for each of the 3 analyses.

Methods and Results

Classification of stressor and response indicators

The WSA partitioned the continuous gradient of each continuous stressor and biological response indicator into 3 condition classes. Good, fair, and poor classes were defined to represent indicator ranges that are, respectively, not different from, somewhat different from, and markedly different from the range sampled at a set of independently determined, least-disturbed reference sites (Stoddard et al. 2006b, Van Sickle et al. 2006).

Condition class thresholds for the WSA (Table 1)

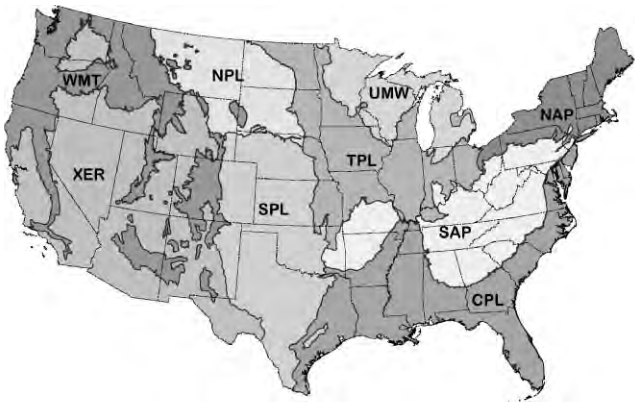

FIG. 1. Aggregated ecoregions (Omernik 1987) in the US used for assigning region-specific condition class thresholds (Table 1). CPL = Coastal Plains, NAP = Northern Appalachians, NPL = Northern Plains, SAP = Southern Appalachians, SPL = Southern Plains, TPL = Temperate Plains, UMW = Upper Midwest, WMT = Western Mountains, and XER = Xeric.

were assigned separately for each of 9 aggregated ecoregions (Fig. 1; Omernik 1987). Most thresholds were based on regional distributions of observed values at reference sites. For example, thresholds for total N (N), total P (P), and salinity (SAL) were equated to the estimated 25^{th} percentile (good/fair) and 5^{th} percentile (fair/poor) of each indicator observed in the nutrient reference sites in each region. A similar approach was used for the MIBI (Herlihy et al. 2008) and for the 4 physical-habitat stressors. In some cases, WSA thresholds varied widely across the 9 ecoregions (Table 1). Thus, the good, fair, or poor classification of any WSA site should be interpreted relative to the least-disturbed reference conditions of the region in which that site is located.

We assessed stressors and responses in terms of condition classes rather than continuous indicators for 3 reasons. First, condition classes provide a common currency for comparing stressors that are not commensurate in their original units (Achen 1982, Greenland et al. 1986). Second, continuous indicators such as instream fish habitat cover (I-SFH) and the MIBI have measurement scales that are unfamiliar to managers, policymakers, and the general public (Van Sickle et al. 2006). Third, the use of classes allowed us to express stressor–response associations in terms of *risk*, a concept that is familiar to most people.

Relative extent and relative risk

Definitions and properties.—Let S_P and S_G denote the event that a stressor, S, is in poor or good condition, respectively. Then the *relative extent* of S_P is the proportion of total regional stream length in poor condition for that stressor (Van Sickle et al. 2006). Relative extent is equivalent to the probability of finding S_P in a randomly selected stream, denoted by $Pr(S_P)$. $Pr(Y_P)$ denotes the relative extent of poor condition for a biological response indicator, Y.

Relative risk (RR) is a ratio of 2 probabilities (Lachin 2000). Its numerator is $Pr(Y_P|S_P)$, defined as the conditional probability (i.e., the risk) of finding poor biological condition (Y_P) at a site, given that the site is also in poor condition for the stressor. The denominator is the baseline conditional probability of Y_P, given that the stressor is in good condition at the site, i.e.,

$$RR = \frac{Pr(Y_P|S_P)}{Pr(Y_P|S_G)}. \quad [1]$$

RR measures the strength of association between the condition classes of a biological response indicator (Y) and a single stressor indicator (S). If $RR = 1.0$, then poor biological condition is equally likely to occur at a site regardless of whether its stressor condition is poor or good. Thus, $RR = 1.0$ indicates no association between the stressor and Y. Values of $RR > 1.0$ measure the increased risk that Y_P occurs whenever S_P occurs.

Methods: application of RR and extent to the WSA.—The WSA assessed 4 stream-chemistry stressors: N, P, SAL, and acidification. It also assessed 4 physical-habitat indicators of stress: riparian disturbance (RD), streambed sediments (SS), I-SFH, and riparian vegetation cover (RVC). The WSA estimated the extent of poor condition of each stressor and also of 2 biological response indicators: MIBI and an observed/expected (O/E) ratio measuring macroinvertebrate taxon loss (USEPA 2006, Paulsen et al. 2008). We estimated RR of each stressor (except acidification) for each response indicator. We did not estimate acidification risks because only 10 of the 1390 sampled WSA sites were classified as poor for this stressor. For simplicity, we demonstrate the estimation of stressor risks only for MIBI.

In the WSA, extent and RR were estimated as described by Van Sickle et al. (2006). Stream sampling sites were selected with unequal probabilities, based on a sampling weight for each site that equaled the total length of stream of that type within the overall stream population (Stehman and Overton 1994, Lohr 1999, Herlihy et al. 2000). Stressor extent was then estimated as the sum of sampling weights for sites in S_P condition, divided by the total sum of weights for all selected sites. This total included some selected sites (4.1% of total length, nationwide) that could not be sampled, primarily because of physical inaccessibility or access denials from landowners (USEPA 2006).

The WSA estimated RR for a stressor from a 2×2

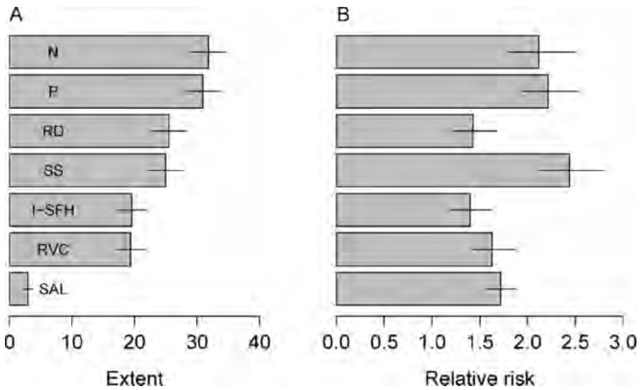

FIG. 2. Estimated extents (% total stream length) of poor condition for 7 stressors (A), and their relative risk for macroinvertebrate index of biotic integrity (B), in wadeable streams at the US national scale (lower 48 states). Relative risk contrasts poor with good conditions, and it excludes sites in fair condition. Stressor codes: N = total N, P = total P, RD = riparian disturbance, SS = streambed sediments, I-SFH = instream fish habitat, RVC = riparian vegetation cover, SAL = salinity. Bars are 95% confidence intervals.

contingency table containing summed weights for 4 subsets of sites in combinations of Y_P or Y_G and S_P or S_G condition (Appendix 1). All sites in the fair class for either Y or S were excluded from these sums. Thus, the *RR*s reported for WSA (USEPA 2006, Paulsen et al. 2008) express the association between 2 extreme classes of stressor and response condition, an approach also used in earlier stream surveys (Stoddard et al. 2005, 2006a).

We report normal-theory 95% confidence intervals for *RR* and extent, based on their estimated standard errors (Van Sickle et al. 2006). We interpret confidence intervals for the family of 7 stressors conservatively, rather than attempting formal adjustments to maintain a familywise confidence level (Miller 1981). We used Taylor linearization to estimate the standard errors of log(*RR*) and extent from the estimation variances of cell and marginal totals in Appendix 1 (Van Sickle et al. 2006). We also used the local-neighborhood variance estimator of Stevens and Olsen (2003), applied to the spatially balanced WSA sampling design (Stevens and Olsen 2004), to minimize cell and marginal variances. Free software for local neighborhood variances and *RR* estimates is available at www.epa.gov/nheerl/arm/.

Results: application of RR *and extents for the WSA.*—Viewed side by side, estimates of stressor extent and *RR* express the prevalence and effect size dimensions of concern about stressors (Fig. 2A, B; Paulsen et al. 2008). Each stressor's confidence interval on *RR* excluded 1.0 (no association; Fig. 2B), a result indicating that every stressor had a significant association with MIBI condition at the national scale (lower 48 states). National-scale *RR*s for the N, P, and SS stressors were >2.0, and these stressors had the greatest extent of poor condition, >25% of total stream length in each case. The combination of greatest extents and highest *RR* for MIBI suggests that N, P, and SS are of greatest nationwide concern. At the other extreme, the poor-SAL condition had the 4^{th} largest relative risk for MIBI, but it occurred in fewer than 3% of streams, suggesting that SAL is of little concern when viewed from a national perspective.

Extent and *RR* are complementary measures of stressor concern (Van Sickle et al. 2006). However, they might not fully satisfy regional assessment needs for 2 reasons. First, it is not clear how best to combine the 2 measures into a single overall assessment when their individual assessments are at odds. For example, RD in the WSA had a greater extent, but a smaller *RR*, than did RVC (Fig. 2A, B). Second, *RR* expresses the stressor–response association at the site scale because it is the increased likelihood of Y_P occurring at a site, given that S_P occurs at a site. Thus, *RR* by itself does not measure a stressor's overall potential effect on a regional stream population. Any such measure would have to account for stressor prevalence in the population, as well as its effect on individual streams (Lachin 2000).

These arguments suggest the need for another index of stressor importance, one that combines a stressor's prevalence (extent) and its site-scale potential effect (*RR*) into a single measure of its overall effect on the stream population. The *population attributable risk* is such an index.

*Population attributable risk (*AR*)*

Definitions and properties.—*AR* has been used widely in epidemiology to describe the contribution of a risk factor to the prevalence of disease in a population (Gefeller 1992, Uter and Pfahlberg 1999). As shown below, *AR* can be expressed as a combination of *RR* and stressor extent. This combination will make sense only if Y and S are each defined as dichotomous condition classes. However, the WSA and earlier assessments did not use consistent classes for *RR* and extent (Stoddard et al. 2005, 2006a, Van Sickle et al. 2006). They defined *RR* as a comparison between the 2 extremes of stressor and response condition (poor vs good, ignoring fair sites), whereas extent was defined as the proportion of poor streams relative to the total for all 3 classes (poor, fair, and good).

To create 2 exhaustive, mutually exclusive classes for each S and Y indicator, we combine the fair and good classes into a single class of *not-poor* condition (N). All subsequent definitions and applications of *AR*, *RR*, and extent will contrast the poor and not-poor condition classes of Y and of each stressor. This example of combining classes differs from the approach used historically by the Environmental Protection Agency (EPA) Office of Water. That office has chosen to report water quality as either "fully supporting" or "not fully supporting" aquatic life use, where the "not fully supporting" class could be formed by combining "partially supporting" and "not supporting" (analogous to our fair and poor) into a single class.

AR *assumptions.—AR* is derived directly from 3 assumptions about the relationships between stressors and a biological response indicator. First, *AR* explicitly assumes stressor *causality*. That is, the poor condition of a stressor at a site is assumed to have caused that site to have an increased probability of poor biological condition. Second, *AR* assumes that a stressor's causal effects would be completely *reversed* if that stressor were eliminated. In the context of our condition classes, a stressor would be eliminated from a regional population of streams if every stream in poor stressor condition were converted to not-poor condition through restoration or remediation activities. Third, *AR*, like *RR*, assumes that the effects of multiple stressors are *independent* enough that each stressor's effect can be estimated reliably in isolation from other stressors. In the Discussion, we critically evaluate these 3 assumptions in the context of regional assessments of aquatic stressors. We also describe an application of *AR* to groups of nonindependent stressors.

AR *derivation.*—Recall that $[Pr(Y_P)]$ is the probability of finding poor biological condition in a stream selected at random from a regional population. Suppose that stressor S could be eliminated from the population, so that all streams currently in poor stressor condition (S_P) could be converted to not-poor condition (S_N). Under the reversibility assumption of *AR*, this blanket conversion would cause the probability of poor biological condition in every formerly stressed stream to decline from its stressed level, $Pr(Y_P|S_P)$, to the unstressed level, $Pr(Y_P|S_N)$. As a result, stressor elimination would lead to all streams in the population having a probability of poor biological condition equal to $Pr(Y_P|S_N)$. Attributable risk expresses the amount of decrease in the overall probability of poor biological condition, from $Pr(Y_P)$ to $Pr(Y_P|S_N)$, as a proportion of the original probability, $Pr(Y_P)$.

Thus, we define the *AR* of a stressor to be the proportional reduction in the extent of poor biological condition that would be achieved if that stressor were eliminated from the population (Walter 1976, Uter and Pfahlberg 1999, Lachin 2000):

$$AR = \frac{Pr(Y_P) - Pr(Y_P|S_N)}{Pr(Y_P)}. \qquad [2]$$

After some algebra (see Appendix 2), *AR* can be expressed as a combination of a stressor's extent and its *RR* (Uter and Pfahlberg 1999, Lachin 2000):

$$AR = \frac{Pr(S_P)(RR - 1)}{1 + Pr(S_P)(RR - 1)}. \qquad [3]$$

Equation 3 shows that a stressor has 0 *AR* if it has either 0 extent $[Pr(S_P) = 0]$ or no association with the biological response indicator ($RR = 1$). Also, the equation shows that, if $AR \neq 0$, then any increase in either a stressor's extent or its *RR* also will increase its *AR*. In applications, we report $100 \times AR$ as the % reduction in $Pr(Y_P)$ that could be achieved by eliminating a stressor.

Methods: application of AR *for the WSA.*—We used the MIBI as the response indicator (Y) and estimated *AR* for each of the 7 WSA stressors. We also re-estimated *RR* and extent values for each stressor based on the new dichotomous classes of poor vs not-poor condition.

We estimated *AR* from the same 2×2 table used to estimate extent and *RR* (Appendix 1). Our estimates included only those WSA sites ($n = 1352$) with complete data for MIBI and all 7 stressors. We made estimates at the US national level (contiguous 48 states) and for each of 3 major regions (Paulsen et al. 2008; Table 1): Eastern Highlands ($n = 265$ sampled sites), Plains and Lowlands ($n = 401$), and West ($n = 686$).

We modified the Taylor-linearization approach of Van Sickle et al. (2006) to construct confidence intervals (CIs) for *AR*. The *AR* estimate (Appendix 1) can be rewritten as $AR_{est} = 1 - G$, where $G = (c/[a+c])/([c+d]/[a+b+c+d])$ (see Appendix 1 for definitions of a, b, c, and d). We first estimated the standard error for $\log(G) = \log(c) - \log(a+c) - \log(c+d) + \log(a+b+c+d)$ by applying Taylor linearization and local variance estimation (Stevens and Olsen 2003) to each term on the right-hand side (see equation A2 and subsequent text in Van Sickle et al. 2006). We then constructed a normal-theory CI for $log(G)$ based on its standard error. Back-transformation of the endpoints of this CI gave the CI for G, whose endpoints were transformed again to give the CI for $AR_{est} = 1 - G$ (Fleiss 1979, Liu 2001).

Results: application of AR *for the WSA.*—Side-by-side estimates show how *AR* expresses the combined magnitudes of extent and *RR* (Fig. 3A–D). For example, in the West, N, P, and SS stressors had similar extents (Fig. 3D). However, SS had the highest *RR* of the 3 stressors and hence had the greatest *AR*. In

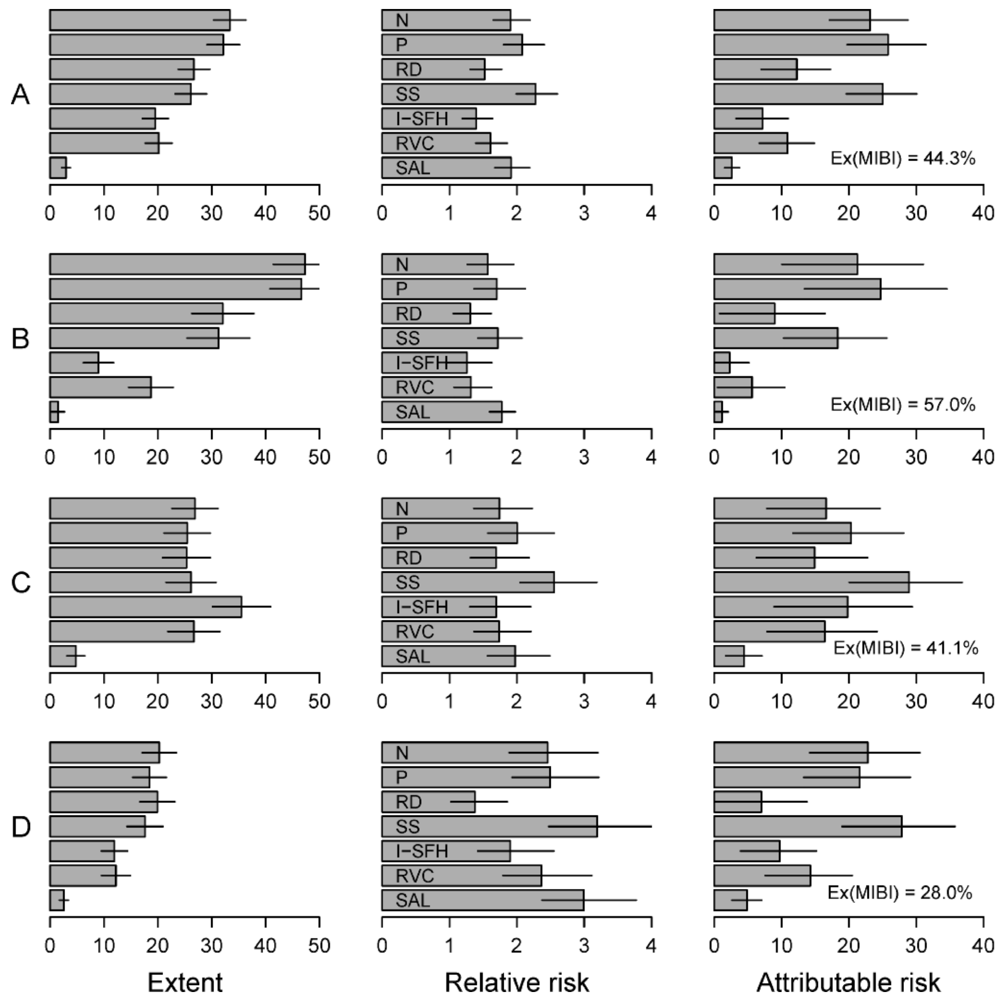

FIG. 3. Estimated extent (% total stream length) of poor condition for 7 stressors, and their relative risk and attributable risk for the macroinvertebrate index of biotic integrity (MIBI) in the US (national = contiguous 48 states) (A), and in the Eastern Highlands (B), Plains and Lowlands (C), and West (D) regions. Attributable risk is expressed as the % reduction in the extent of poor MIBI condition (see derivation in Methods and Results). Risk estimates contrast poor with not-poor conditions. Ex(MIBI) is the extent of poor condition for MIBI. Stressor codes: N = total N, P = total P, RD = riparian disturbance, SS = streambed sediments, I-SFH =← instream fish habitat, RVC = riparian vegetation cover, SAL = salinity. Bars are 95% confidence intervals.

the Eastern Highlands, this pattern was reversed (Fig. 3B). The similar *RR*s for all 3 stressors combined with greater extents of N and P resulted in SS having the lowest *AR* of the 3 stressors. *RR* for SAL was close to those of N, P, and SS. However, the *AR* for SAL was the lowest (<5%) of all stressors in every region because of SAL's small extent.

N, P, and SS had the greatest *AR*s of all 7 stressors, both nationally and regionally (Fig. 3A–D). For example, in the Eastern Highlands, *AR*s for N, P, and SS were 21%, 25%, and 18%, respectively (Fig. 3B). If any one of these stressors could be eliminated, equation 2 predicts that the extent of poor MIBI condition would be reduced from 57% (Fig. 3B) to (1 – *AR*/100) × 57%, i.e., to 45%, 43%, or 47% for N, P, and SS, respectively, under the assumption of complete reversibility of stressor effects. Similar predictions can be made for other stressors and regions, given the estimated extents of poor MIBI. For example, stressors have similar *AR*s in the West as seen elsewhere, but their elimination would produce smaller absolute decreases in the West's extent of poor MIBI (28.0%) (Fig. 3D) because it is already lowest among the 3 regions.

Combined AR *for multiple correlated stressors*

Definitions and properties.—AR, like *RR*, assumes that each stressor's effect is independent of the others and can be estimated in isolation from other stressors. Thus, *AR* and *RR* are analogous to bivariate correlations as models for the relationship between stressor and response indicators. In addition, the definition of *AR* assumes that one stressor could be eliminated from a regional population without altering the distributions of any other stressors. These assumptions are unrealistic, to the degree that stressors are highly correlated. For example, within the WSA reference sites, continuous variables representing human disturbance were sufficiently correlated that their 1st principal component could adequately represent a single, generalized disturbance gradient (Herlihy et al. 2008).

We suggest estimating the *combined attributable risk* for highly correlated stressors. The combined *AR* of a group of 2 stressors is the proportional reduction in $Pr(Y_P)$ that would be achieved by eliminating that group of stressors (Land et al. 2001). We calculated the combined *AR* from equation 2, with S_N replaced by the joint occurrence of condition *N* for all stressors in the group. Thus, our combined *AR* defines a site to be poor for the group of stressors if at least one of the stressors in the group is in poor condition.

The *AR*s of individual stressors sum to their combined *AR* only if the stressors are uncorrelated (Walter 1980, Land et al. 2001). For example, the 7 WSA stressors have a combined *AR* of 68% at the national level, but their individual *AR*s (Fig. 3A) sum to 107% because of stressor intercorrelations. Thus, the *AR* estimates of Fig. 3A overestimate the effects of individual stressors because *AR* of each stressor is expressing some portion of the effects from other correlated stressors.

Depending on the strengths of between-stressor correlations, more reliable estimates of stressor effects can be made by comparing the combined *AR*s of separate groups of closely related stressors. Stressors might be grouped based on one or more of the following criteria: 1) stressors within a group represent closely linked physical, chemical, or biological processes that affect the response indicator, 2) stressors within the same group are more strongly statistically correlated than are stressors from different groups, and 3) restoration or remediation efforts would be more likely to manage (i.e., reduce or eliminate) all stressors within a group than to manage them separately.

Methods: application of combined AR *for the WSA.*—To illustrate, we compare the combined *AR*s of 2 groups each containing 2 WSA stressors. First, we combined N and P stressors into a "nutrients" group. These stressors both affect aquatic systems through the enrichment of primary production, and they probably would be managed together through reduction of fertilization and wastewater inputs from watersheds (USEPA 2006). Second, we combined I-SFH and RVC stressors into a "woody vegetation" group. Low RVC can lead to increased water temperatures and reduced ability to intercept nutrients and sediments from nonpoint sources. Low RVC also leads to reduced large wood in streams, fewer streambank roots, and less overhanging vegetation, all of which are important components I-SFH. RVC and I-SFH stressors would both be reduced by managing for enhanced riparian woody vegetation.

The nationwide product-moment correlations between categorical stressors (Zar 1999, Van Sickle et al. 2006) are 0.41 for N vs P and 0.20 for I-FSH vs RVC. These within-group stressor correlations are both larger than the 4 between-group correlations, which range from 0.02 to 0.12. Thus, our 2 groups of 2 correlated stressors each should have little confounding of group-level effects.

Results: application of the combined AR for the WSA.—Based on point estimates at the national level and in 2 of 3 regions, nutrient stressors had a greater combined *AR* for MIBI than did woody vegetation stressors (Fig. 4). However, in the Plains and Lowlands, woody vegetation had a slightly higher combined *AR* than did nutrients, probably because the extent of poor conditions was greater for the 2 woody vegetation stressors than for nutrients. However, CIs on combined *AR* were relatively wide, resulting in overlap within the Plains and Lowlands and West (Fig. 4).

The combined *AR* of all 7 WSA stressors was 68%. If one assumes reversibility of stressor effects, then this combined *AR* predicts that elimination of all 7 stressors would reduce the national extent of poor MIBI from 44.3% (Fig. 3A) to $(1 - AR/100) \times 44.3\% =$← 14.2%. Thus, elimination of the 7 stressors would not lead to complete elimination of poor MIBI conditions. The combined *AR* of the 7 stressors is <100%, in part, because MIBI is also influenced by natural factors, as well as additional stressors, such as stream temperature, toxic materials, and hydrologic modification, that were not assessed by the WSA.

Discussion

We think that population *AR* is a valuable addition to the extent and *RR* measures for assessing the relative effects of aquatic stressors at regional and national scales. Extent measures the regional prevalence of a stressor, and *RR* measures its site-scale effect. *AR* combines these 2 measures into a new estimate of a stressor's probable population-scale effect. This feature is especially helpful for stressors whose *RR* and extent estimates give disparate pictures of stressor influence, such as seen for SAL (Fig. 2A, B).

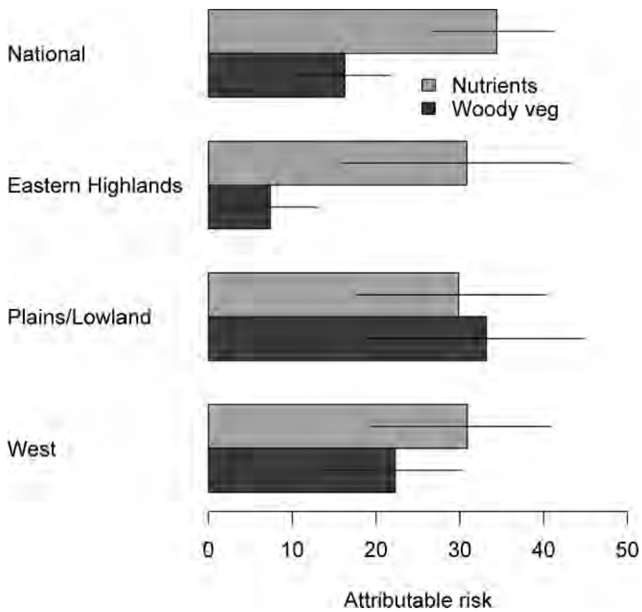

FIG. 4. Regional combined attributable risks of the nutrients and woody vegetation (woody veg) stressor groups for the macroinvertebrate index of biotic integrity (MIBI). Attributable risk is expressed as the % reduction in the extent of poor MIBI condition. Solid lines denote 95% confidence intervals. Bars are 95% confidence intervals.

Assumptions and interpretations of AR

Users of *AR* must contend with and communicate clearly its assumptions of causality and reversibility. We discuss these assumptions here, and they have been examined critically in the literature on human epidemiology (Greenland and Robins 1988, Rockhill et al. 1998, Uter and Pfahlberg 1999).

Causality.—A large *AR* does provide supporting evidence for a presumed causal relationship between stressor and response. Unfortunately, *AR*, like any other correlational statistic, cannot demonstrate causality on its own (Beyers 1998, Shipley 2000, USEPA 2000). The WSA report (USEPA 2006) sketches another line of evidence for causality, namely, the chemical, physical, and biological processes by which each of our 7 stressors could have caused poor MIBI condition. However, a strong argument that a stressor has actually caused impairment requires multiple additional lines of evidence (Suter 1993, USEPA 2000, Collier and Adams 2003). Applications of *AR* to regional assessments should regard stressor causality as an unproven assumption because of the difficulty of assembling multiple lines of causal evidence at a regional scale.

However, causality is not an unreasonable assumption to make for purposes of evaluating stressors at a regional scale. The assumption of causality underlies most observational studies of stressor–response relationships, even though researchers are careful to state that their models are strictly associational. Such studies often use causal language, and they are motivated by the implicit causal assumption that stressors "stress" aquatic systems. For example, Van Sickle et al. (2006) estimated *RR* from co-occurrences of poor stressor and response conditions, while interpreting it as a measure of the severity of stressor effects (see also Yuan and Norton 2003). Yuan and Norton (2004), Kapo and Burton (2006), and de Zwart et al. (2006) also have proposed population-level measures of stressor effects that, like *AR*, are based on associational data and models but assume causality, either implicitly or explicitly. However, unlike AR, these 3 methods make their population estimates by summing the site-specific effect contributions of each stressor across multiple sites in the stream or lake population.

Reversibility.—Potential users of *AR* might be willing to assume causality, but they might balk at assuming reversibility. Even if a stressor could be eliminated, the resulting degree of ecosystem recovery and its timing generally are difficult to predict (Suter 1993). Thus, *AR* estimates are probably best interpreted as upper bounds or best-case scenarios on the benefits that one might expect from eliminating stressors because they assume that stressor effects are completely reversed. For example, in the Eastern Highlands, *AR* predicts that elimination of poor woody vegetation conditions would improve the MIBI condition class in <10% of streams (Fig. 4). Likewise, in the West, elimination of the single stressor with the greatest *AR* (SS) could decrease the extent of poor MIBI by only 8 percentage points (Fig. 3D; equation 2), primarily because that extent is already relatively low (28%). In both cases, *AR* predicts that relatively little regional benefit would accrue from stressor elimination, even under the idealized scenario of full reversibility of stressor effects.

To be minimally realistic, any *AR* scenario must be based on attainable targets for stressor elimination and biological recovery (Rockhill et al. 1998). For example, our WSA application of *AR* used poor and not-poor classes. Application of *AR* to these classes assumes that streams in poor stressor or poor biological conditions could attain at least fair condition (i.e., not poor), following stressor elimination and recovery. Fair condition would seem to be a realistic target. In contrast, application of *AR* to good (reference) vs not-good conditions would have to assume that every stream could be converted to reference conditions through stressor eliminations. This scenario could be much less realistic, particularly in regions where reference conditions are defined primarily by near-pristine streams.

Condition classes for AR

AR requires that continuous stressor and biological response gradients be converted to dichotomous condition classes, a process that collapses gradients much more severely, for example, than do the 6 general classes of aquatic life use condition identified by Davies and Jackson (2006). Dichotomous classifications might be difficult to accept, in part because all data analysis outcomes for such classes depend critically on the single boundary chosen to separate them. In addition, one would expect the average conditions of dichotomous classes to differ less than the average conditions of the extremes of 3 classes. Thus, nearly all of the *RR*s of Fig. 3A (based on 2 classes) are lower than corresponding values based on the extremes of 3 classes (Fig. 2B). Use of dichotomous condition classes might lead to some concerns, but in practice, states and tribes routinely set standards and criteria to define waters as being impaired or not impaired.

AR *for correlated stressors*

In practice, managers often might want to compare the combined *AR*s of groups of closely related correlated stressors (Fig. 4) rather than to estimate their approximate individual effects. Van Sickle et al. (2006) also combined highly correlated stressors, in a qualitative way, to help interpret their relative risks. We advocate a combined-stressor approach for its simplicity and for its realistic emphasis on joint modeling and management of stressors whose effects are confounded. Future research might help develop additional methods for combining class-based stressor variables.

An alternative strategy for multiple stressors is to adjust the *AR* of an individual stressor for the effects of other correlated stressors or nonstressor covariates (Walter 1980, Eide and Gefeller 1995, Benichou 2001). The adjusted *AR* is analogous to a regression coefficient for a stressor in a multiple regression model that includes other stressors. The adjusted *AR* might be especially helpful in dealing with the potentially confounding effects of natural (nonanthropogenic) variables on biological responses (Herlihy et al. 2008).

However, *AR* adjustments for multiple dichotomous stressors, as in the WSA case, are complex and might require additional assumptions and models (Benichou 2001). In addition, adjusted *AR*s (like multiple regression models), cannot partition unambiguously the overall variance explained by multiple correlated stressors into their individual contributions. A "fair" partitioning of joint stressor effects has been suggested for adjusted *AR* (Cox 1985, Eide and Gefeller 1995, Land et al. 2001) and for multiple regression (Kruskal 1987, Grömping 2007). However, the partitioning algorithm is ad hoc, derived neither from statistical theory nor process-based knowledge of joint effects (Cox 1985). We have not illustrated *AR* adjustments and partitioning in our paper because of these complications. We anticipate further research on applications of adjusted *AR* to aquatic stressors.

Uncertainty of AR *estimates*

We have stated our results for extent, *RR*, and *AR* without considering the estimation uncertainties of the 3 statistics. The width of our *AR* CIs, relative to *AR* estimates, are greater than the corresponding relative widths for extent and *RR* CIs, and the *AR* intervals for different stressors often overlap (Figs 3, 4). The wide CIs for *AR* seem surprising, given the large samples from which *AR* was estimated (range: 265 [Eastern Highlands] to 1352 [nationwide]). However, other methods for estimating *AR* standard errors (Liu 2001, Lehnert-Batar et al. 2006), including jackknifing, yielded even wider CIs for our WSA *AR* estimates. We do not know the exact reasons for the relatively high uncertainty in *AR* estimates. Note, however, that the *AR* estimation uncertainty compounds the estimation uncertainties of *RR* and stressor extent (equation 3).

Sampling designs for estimating AR *and* RR

AR and *RR* can be estimated directly only from the cross-sectional sampling design used by the WSA and similar regional aquatic surveys (Stoddard et al. 2005, 2006a, USEPA 2006). In this design, sampling units (stream sites) represent a cross-section of the regional stream population and are selected at random without regard for their stressor or biological condition status (Suter 1993, Lachin 2000). The cross-sectional design yields unbiased estimates of each cell in Appendix 1 and thus unbiased estimates of extent, *RR*, and *AR* (Walter 1976). Some cross-sectional designs use equal-probability weighting (simple random sampling) when selecting sites, rather than the unequal weighting used by recent surveys (Stoddard et al. 2005, 2006a, USEPA 2006). In the equal-weighting case, risk and extent estimates can be calculated from site counts rather than weight sums (Appendix 1).

RR also could be estimated from a 2-group gradient study, in which one first selects an approximately equal number of streams in poor condition (group 1) and in not-poor condition (group 2) for a given S.

Estimates of Y_P, made separately within each group, then give the numerator and denominator of equation 1. This 2-group gradient study design is analogous to a prospective design in human health research, in which exposed (to a risk factor) and unexposed subjects are followed over time to estimate their probabilities of developing a disease. Like the prospective design, the 2-group gradient study cannot, by itself, estimate *AR* because *AR* also requires knowledge of $Pr(S_P)$ (equation 3; Walter 1976). Retrospective sampling, another design commonly used in human health studies, would select groups of streams based on their biological response condition rather than their stressor condition and then estimate stressor extent within each group. However, this 3rd design can estimate neither *RR* nor *AR* without additional information (Walter 1976).

Final remarks

AR estimates the expected regional-scale improvement in aquatic biological condition that would result from regional elimination of a stressor, under the assumptions that the stressor caused impairment and that its effects can be reversed. At a regional scale, the *AR* assumptions (causality and reversibility of stressor effects) cannot be validated realistically and must be viewed as "what if" scenarios. However, we think that such scenarios can help policymakers and managers to identify key regional stressors and to estimate the possible benefits of stressor remediation. When used together with *RR* and extent, *AR* is a potentially useful tool for regional and national-scale assessment of aquatic stressors.

Acknowledgements

We thank Chuck Hawkins, Susan Norton, Hal Walker, and an anonymous referee for their ideas and critical readings of the manuscript. We also thank Ulrike Grömping for guidance with adjusted attributable risks. This research was funded by the US EPA. The manuscript has been subjected to the agency's peer and administrative review and approved for publication as an EPA document. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

Literature Cited

ACHEN, C. H. 1982. Interpreting and using regression. Sage Publications, Beverly Hills, California.

BENICHOU, J. 2001. A review of adjusted estimators of attributable risk. Statistical Methods in Medical Research 10:195–216.

BEYERS, D. W. 1998. Causal inference in environmental impact studies. Journal of the North American Benthological Society 17:367–373.

COLLIER, T. K., AND S. M. ADAMS. 2003. Establishing causal relationships between environmental stressors and biological effects in field studies. Human and Ecological Risk Assessment 9:15.

COX, L. A. 1985. A new measure of attributable risk for public health applications. Management Science 31:800–813.

DAVIES, S. P., AND S. K. JACKSON. 2006. The biological condition gradient: a descriptive model for interpreting change in aquatic ecosystems. Ecological Applications 16:1251–1266.

DE ZWART, D., S. D. DYER, L. POSTHUMA, AND C. P. HAWKINS. 2006. Predictive models attribute effects on fish assemblages to toxicity and habitat alteration. Ecological Applications 16:1295–1310.

EIDE, G. E., AND O. GEFELLER. 1995. Sequential and average attributable fractions as aids in the selection of prevention strategies. Journal of Clinical Epidemiology 48:645–655.

FLEISS, J. L. 1979. Inference about population attributable risk from cross-sectional studies. American Journal of Epidemiology 110:103–104.

GEFELLER, O. 1992. An annotated bibliography on the attributable risk. Biometrical Journal 34:1007–1012.

GREENLAND, S., J. J. SCHLESSELEMAN, AND M. H. CRIQUI. 1986. The fallacy of employing standardized regression coefficients and correlations as measures of effect. American Journal of Epidemiology 123:203–208.

GREENLAND, S., AND J. M. ROBINS. 1988. Conceptual problems in the definition and interpretation of attributable fractions. American Journal of Epidemiology 128:1185–1197.

GRÖMPING, U. 2007. Estimators of relative importance in linear regression based on variance decomposition. American Statistician 61:139–147.

HERLIHY, A. T., D. P. LARSEN, S. G. PAULSEN, N. S. URQUHART, AND B. J. ROSENBAUM. 2000. Designing a spatially-balanced, randomized site selection process for regional stream surveys: the EMAP Mid-Atlantic Pilot Study. Environmental Monitoring and Assessment 63:95–113.

HERLIHY, A. T., S. G. PAULSEN, J. VAN SICKLE, J. L. STODDARD, C. P. HAWKINS, AND L. L. YUAN. 2008. Striving for consistency in a national assessment: the challenges of applying a reference condition approach at a continental scale. Journal of the North American Benthological Society 27:860–877.

KAPO, K. E., AND G. A. BURTON. 2006. A geographic information systems-based weights-of-evidence approach for diagnosing aquatic ecosystem impairment. Environmental Toxicology and Chemistry 25:2237–2249.

KARR, J. R. 1981. Assessment of biotic integrity using fish communities. Fisheries 6:21–27.

KRUSKAL, W. 1987. Relative importance by averaging over ordering. American Statistician 41:6–10.

LACHIN, J. M. 2000. Biostatistical methods: the assessment of relative risk. John Wiley and Sons, New York.

LAND, M., C. VOGEL, AND O. GEFELLER. 2001. Partitioning methods for multifactorial risk attribution. Statistical Methods in Medical Research 10:217–230.

LEHNERT-BATAR, A., A. PFAHLBERG, AND O. GEFELLER. 2006. Comparison of confidence intervals for adjusted attributable risk estimates under multinomial sampling. Biometrical Journal 48:805–819.

LIU, K.-J. 2001. Notes on interval estimation of the attributable risk in cross-sectional sampling. Statistics in Medicine 20:1797–1809.

LOHR, S. L. 1999. Sampling: design and analysis. Brooks/Cole, Pacific Grove, California.

MILLER, R. G. 1981. Simultaneous statistical inference. 2nd edition. Springer-Verlag, New York.

OMERNIK, J. M. 1987. Ecoregions of the coterminous United States. Annals of the Association of American Geographers 77:118–125.

PAULSEN, S. G., A. MAYIO, D. V. PECK, J. L. STODDARD, E. TARQUINO, S. M. HOLDSWORTH, J. VAN SICKLE, L. L. YUAN, C. P. HAWKINS, A. T. HERLIHY, P. R. KAUFMANN, M. T. BARBOUR, D. P. LARSEN, AND A. R. OLSEN. 2008. Condition of stream ecosystems in the US: an overview of the first national assessment. Journal of the North American Benthological Society 27:812–821.

RICE, J. A. 1988. Mathematical statistics and data analysis. Wadsworth and Brooks/Cole, Pacific Grove, California.

ROCKHILL, B., B. NEWMAN, AND C. WEINBERG. 1998. Use and misuse of population attributable fractions. American Journal of Public Health 88:15–19.

SHIPLEY, B. 2000. Cause and correlation in biology. Cambridge University Press, Cambridge, UK.

STEHMAN, S. V., AND W. S. OVERTON. 1994. Environmental sampling and monitoring. Pages 263–306 *in* G. P. Patil and C. R. Rao (editors). Handbook of statistics. Volume 12. Elsevier, New York.

STEVENS, D. L., AND A. R. OLSEN. 2003. Variance estimation for spatially-balanced samples of environmental resources. Environmetrics 14:593–610.

STEVENS, D. L., AND A. R. OLSEN. 2004. Spatially balanced sampling of natural resources. Journal of the American Statistical Association 99:262–278.

STODDARD, J. L., A. T. HERLIHY, B. H. HILL, R. M. HUGHES, P. R. KAUFMANN, D. J. KLEMM, J. M. LAZORCHAK, F. H. MCCORMICK, D. V. PECK, S. G. PAULSEN, A. R. OLSEN, D. P. LARSEN, J. VAN SICKLE, AND T. R. WHITTIER. 2006a. Mid-Atlantic Integrated Assessment (MAIA)—state of the flowing waters report. EPA/620/R-06/001. Office of Research and Development, US Environmental Protection Agency, Washington, DC.

STODDARD, J. L., D. P. LARSEN, C. P. HAWKINS, R. K. JOHNSON, AND R. H. NORRIS. 2006b. Setting expectations for the ecological condition of streams: the concept of reference condition. Ecological Applications 16:1267–1276.

STODDARD, J. L., D. V. PECK, S. G. PAULSEN, J. VAN SICKLE, C. P. HAWKINS, A. T. HERLIHY, R. M. HUGHES, P. R. KAUFMANN, D. P. LARSEN, G. LOMNICKY, A. R. OLSEN, S. A. PETERSON, P. L. RINGOLD, AND T. R. WHITTIER. 2005. An ecological assessment of western streams and rivers. EPA 620/R-05/005. Office of Research and Development, US Environmental Protection Agency, Washington, DC.

SUTER, G. W. 1993. Ecological risk assessment. Lewis Publishers, Chelsea, Michigan.

USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2000. Stressor identification guidance document. EPA-822-B-00–025. Office of Water, US Environmental Protection Agency, Washington, DC.

USEPA (US ENVIRONMENTAL PROTECTION AGENCY). 2006. Wadeable streams assessment: a collaborative survey of the nation's streams. EPA 841/B-06/002. Offices of Research and Development and of Water, US Environmental Protection Agency, Washington, DC.

UTER, W., AND A. PFAHLBERG. 1999. The concept of attributable risk in epidemiological practice. Biometrical Journal 41: 985–993.

VAN SICKLE, J., J. L. STODDARD, S. G. PAULSEN, AND A. R. OLSEN. 2006. Using relative risk to compare the effects of aquatic stressors at a regional scale. Environmental Management 38:1020–1030.

WALTER, S. D. 1976. The estimation and interpretation of attributable risk in health research. Biometrics 32:829–849.

WALTER, S. D. 1980. Prevention for multifactorial diseases. American Journal of Epidemiology 112:409–416.

YUAN, L. L., AND S. B. NORTON. 2003. Comparing responses of macroinvertebrate metrics to increasing stress. Journal of the North American Benthological Society 22:308–322.

YUAN, L. L., AND S. B. NORTON. 2004. Assessing the relative severity of stressors at a watershed scale. Environmental Monitoring and Assessment 98:323–349.

ZAR, J. H. 1999 Biostatistical analysis. 4th edition. Prentice–Hall, Upper Saddle River, New Jersey.

Received: 21 November 2007
Accepted: 17 June 2008

APPENDIX 1. Estimating extent, relative risk (*RR*), and attributable risk (*AR*) from a 2 × 2 table of observed stressor and response occurrences (adapted from Van Sickle et al. 2006). Table entries (*a–d*) are sums of sampling weights (for unequal-probability sampling) or site counts (for simple random sampling) across all sites having the stated combination of stressor (S) and response (Y) class. Stressor and response classes are defined as not-poor (N; good + fair) and poor (P). When estimating relative risk (RR), sites in fair condition for the stressor or response were excluded from the not-poor class (N) to provide a contrast between good and poor conditions (USEPA 2006).

	S	
Y	N	P
N	a	b
P	c	d

Estimates of extent, RR, and AR are given by

$(\text{Extent of } S_P) = (b+d)/(a+b+c+d)$

$(\text{Extent of } Y_P) = Pr(Y_P)_{est} = (c+d)/(a+b+c+d)$

$RR = (d/[b+d])/(c/[a+c])$
$AR = [Pr(Y_P)_{est} - c/(a+c)]/Pr(Y_P)_{est}$

APPENDIX 2. Derivation of equation 3 from the definition of attributable risk (AR) (equation 2).

By the law of total probability (Rice 1988), $Pr(Y_P)$ can be written as

$$Pr(Y_P) = Pr(Y_P|S_P)Pr(S_P) + Pr(Y_P|S_N)Pr(S_N) \qquad [A.1]$$

In equation A.1, substitute $(1 - Pr(S_P))$ for $Pr(S_N)$, and insert the resulting expression for $Pr(Y_P)$ into the numerator of equation 2 to yield

$$AR = \frac{Pr(S_P)[Pr(Y_P|S_P) - Pr(Y_P|S_N)]}{Pr(Y_P)}. \qquad [A.2]$$

In equation A.2, the risk reduction achieved by eliminating the stressor is expressed as the extent of poor stressor condition times the difference in risk of Y_P under poor vs not-poor stressor conditions (Lachin 2000). Note that equation A.2 compares these 2 conditional risks of Y_P as a difference, whereas relative risk (RR) (equation 1) compares the same 2 risks as a ratio.

To complete the derivation of equation 3, substitute equation A.1 for $Pr(Y_P)$ in the denominator of equation A.2, and substitute $(1 - Pr(S_P))$ for $Pr(S_N)$ to give

$$AR = \frac{Pr(S_P)[Pr(Y_P|S_P) - Pr(Y_P|S_N)]}{Pr(Y_P|S_P) + Pr(S_P)[Pr(Y_P|S_P) - Pr(Y_P|S_N)]} \qquad [A.3]$$

Divide the numerator and denominator of equation A.3 by $Pr(Y_P|S_N)$ and apply the definition of relative risk (RR) (equation 1) to obtain equation 3.